

# Genetic Machine Learning Approach for Data Fusion Applications in Dense Wireless Sensor Networks

A. R. Pinto<sup>1</sup>, Benedito Bitencort<sup>1</sup>,  
M. A. R. Dantas<sup>2</sup> and Carlos B. Montez<sup>1</sup>  
Automation and Systems Department<sup>1</sup>  
Computer Science Department<sup>2</sup>  
Federal University of Santa Catarina, Brazil  
{arpinto, benedito, montez}@das.ufsc.br  
mario@inf.ufsc.br

Francisco Vasques  
Mechanical Engineering Department  
University of Porto, Portugal  
vasques@fe.up.pt

## Abstract

*Wireless Sensor Networks (WSN) are being targeted for use in applications like security, resources monitoring and factory automation. However, the reduced available resources raise a lot of technical challenges. Self-organization in WSN is a desirable characteristic that can be achieved by means of data fusion techniques when delivering reliable data to users. In this paper it is proposed a genetic machine learning algorithm (GMLA) approach that makes a trade-off between quality of information and communication efficiency. GMLA is based on genetic algorithms and it can adapt itself dynamically to environment modifications. The main target of the proposed approach is to achieve self-organization in a WSN application with data fusion. Simulations demonstrate that the proposed approach can optimize communication efficiency in a dense WSN.*

## 1. Introduction

Wireless sensor networks (WSN) are composed by several tiny nodes interconnected by wireless media. These nodes have processor, memory, sensors and limited energy. WSN are usually used to monitor inaccessible and dangerous areas. The low dependability of individual nodes has to be overcome through cooperative effort techniques. However, hundreds of nodes deployed in a monitoring area and sensing scalar data present many technical challenges [3]. One challenge being the reliable delivery of data to users. Users are interested in global information instead of individual data [6]. Data fusion approaches are needed to combine data from different sources in order to generate reliable information. Self-organization is another desirable characteristic in WSN. This way, it is possible to manage a dense WSN without human interference [3].

The IEEE 802.15.4 can be considered a *de facto* standard in WSN [10], allowing low-power and low-rate

wireless networks. This standard can address 65,000 nodes. Nevertheless, whenever the network density increases, the efficiency of data utilization decreases, due to collisions and retransmissions [11].

In this paper, it is considered the IEEE 802.15.4 in beaconless mode with a star topology. Figure 1 shows that when network density is increased, communication efficiency decreases. The efficiency is a metric that measures the ratio between sent and received messages. The other metric presented in Figure 1, QoF, represents, roughly, the average number of received message by the base station, periodically. QoF provides a quality measurement. Both metrics are defined in Section 2.

In this paper, we propose the use of a genetic machine learning algorithm (GMLA) that makes a trade-off between quality of information and communication efficiency. GMLA can be used to dynamically adapt systems due to its genetic behaviour [5]. The main goal of the proposed approach is to achieve self-organization in a WSN application with data fusion. Simulations demonstrated that the proposed GMLA can optimize communication efficiency in a dense WSN.

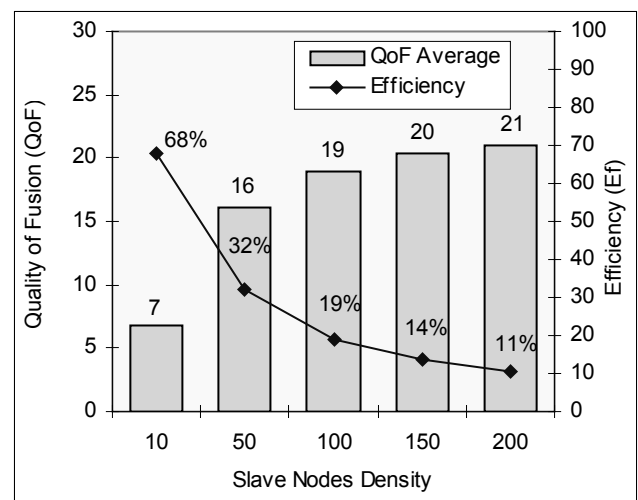
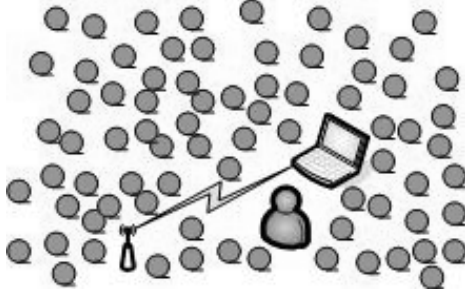


Figure 1: IEEE 802.15.4 behaviour

## 2. Model

The used communication model considers one master node (base station) e  $N$  slave nodes (Figure 2), where the slave nodes periodically sense scalar data. The signal is considered to be homogeneous in the monitoring area. Data collected by slaves are sent to master node that performs data fusion. All the slave nodes reach the master node using only one hop. This way, a parallel data fusion is performed in master node.



**Figure 2: System architecture**

The concept of monitoring session  $S$  is adopted. A monitoring session is a time interval where all slave nodes send periodically sensed data to the master node. A session  $S$  is composed by  $N$  rounds with size  $R$ . Therefore, it is composed by  $0, R, 2R, 3R, \dots, (N-1)R$  rounds. The round concept is used to synchronize nodes, and it also represents the periodicity of the data fusion task. On each round, a slave node can send zero or one message  $M$  containing its sensed data to master node.

All slave nodes are synchronized by the round concept. Each message  $M$  sent by a slave node has an absolute deadline  $D$ , that is the maximum time interval to be delivered to master node. Otherwise, it will not be useful for the data fusion task. This absolute deadline is computed based on a relative deadline  $d$ . We considered an homogeneous architecture where all slave nodes have the same relative deadline. This deadline value is sent by master node in the beginning of the session. The absolute deadline of a slave node at a round  $n$  is  $D=nR+d$ .

The master node performs a data fusion considering just the messages that arrived on time. In this work, the master node just fuses data that arrived in the same round. Therefore, the relative deadline of a message sent in round  $n$  is always  $0 < d < R$ , and consequently, absolute deadline is  $nR < D < (n+1)R$ .

A sending probability ( $S_p$ ) parameter is considered in the model, and all slave nodes have the same  $S_p$ . This parameter controls messages sent by slave nodes within each round. For instance, if  $S_p$  is configured to 0,1, each slave has a probability of 10% to send its message. The signal is considered homogeneous and redundant in the monitoring area, so a well-configured  $S_p$  saves network energy, reducing packets in the WSN.

The sending probability, the round time, and the relative deadline parameters are sent by the master node

in the beginning of each session. Some of these parameters could be valid during all the monitoring session, or they can be changed in a checkpoint  $C$ . Checkpoint is a special round where it is imposed the re-synchronization of parameters based on network condition. Slave nodes do not send messages in a checkpoint round; they just receive new parameter values. The first round of every monitoring session is a checkpoint round and slave nodes always wait for parameter values in this first round.

The master node calculates performance metrics during a checkpoint round in order to tune the WSN. In the proposed model, two metrics are considered: Quality of Fusion ( $Q_{of}$ ) and Efficiency ( $E_f$ ).  $E_f$  is the relation between received messages (messages received by the master node before the deadline) and sent messages and it is calculated according to:

$$E_f = \frac{\sum_{i=1}^N R_m}{E_m} \quad (1)$$

where  $N$  is the number of rounds since previous checkpoint  $C$ ,  $R_m$  is the number of received messages and  $E_m$  is an estimation of the number of messages sent by slave nodes (2). This metric indicates how many messages are used in data fusion task.

$$E_m = S_p \times D_e \times N \quad (2)$$

where  $D_e$  is the density of slave nodes in the WSN deployment. Finally,  $Q_{of}$  is the average number of received messages by master node during a monitoring session, which is evaluated according to:

$$Q_{of} = \frac{\sum_{i=1}^N R_m}{N} \quad (3)$$

The basic idea of the  $Q_{of}$  metric is to represent the quality of information on data fusion. A higher number of messages used in by the data fusion task result in more reliable information.

## 3. Related Work

The proposed model has its roots in previous research works and also in some wireless network standards. The adopted star topology is part of ZigBee technology (based in IEEE 802.15.4 MAC and physical standard). Also, approaches presented in [2,4,8] use star topologies where sensor nodes reach base station in just one hop.

The round concept is showed in [9]. The main goal of this concept is to maximize WSN coverage area, where the monitoring phase (that is equivalent to our session concept) is divided in equal duration rounds.

A metric similar to the proposed  $Q_{of}$  concept is presented in [7], within the context of real-time databases. The metric is called  $Q_{od}$  (Data Base

Freshness) and considers the miss deadline ratio and the data freshness ( $Q_{\odot D}$  levels) as the relevant metrics.

A parallel data fusion scenario is considered in [4], where the master node is not aware of the number of sensor nodes. The data fusion rule (referred as counting rule) imposes that the number of packet must be greater than a defined threshold in order to make a decision.

A serial data fusion based on genetic algorithms is presented in [6]. However, a mobile agent approach for the target detection is used in order to validate a multi-objective genetic algorithm.

#### 4. Genetic Machine Learning Algorithms

The Genetic Machine Learning Algorithm (GMLA), also known as classifier system, is a machine-learning algorithm based on genetic algorithms. These systems can learn syntactically simple rules (classifiers) that guide its performance in an arbitrary environment. A classifier system has three main modules:

- Rule and Message System;
- Apportionment of credit algorithm;
- Genetic Algorithm.

The rule and message system module is a special class of production system. Production system is a computational scheme that employs rules as the main algorithmic goal. Classifiers usually have the following structure:

**if** <condition> **then** <action>

Its meaning is that the predefined action will be taken whenever some predefined condition is satisfied.

##### 4.1. Proposed GMLA

The main target of the proposed approach is to dynamically adapt the sending probability  $S_p$ . In such a way, there is a trade-off between  $Q_{\odot F}$  and  $E_f$ . The set of simulated <condition> and <action> parameters is showed in Table 1.

**Table 1: Classifiers configuration**

Classifier Part	Bits	Meaning
C1	1	0 = decrease 1 = increase
C2	3	000 = [0-12%] 100 = (48-64%] 001 = (12-24%] 101 = (64-72%] 010 = (24-36%] 110 = (72-84%] 011 = (36-48%] 111 = (84-100%)
A1	1	0 = decrease 1 = increase
A2	3	000 = 12% 100 = 64% 001 = 24% 101 = 72% 010 = 36% 110 = 84% 011 = 48% 111 = 100%

According to Table 1, a classifier is composed of four parts: C1, C2, A1 and A2. The form of a classifier is <C1+C2>:<A1+A2>. C1 value indicates if the

efficiency has increased or decreased since the last checkpoint, and C2 indicates the efficiency gain level. A1 value indicates if  $S_p$  will be increased or decreased, and A2 indicates the level of change in  $S_p$ . The overhead imposed by GMLA is much smaller than the overhead of traditional GAs. However, the evolution requires more system execution time. The key of GMLA is that the evolution is done during the execution time, whereas a traditional GA evolves to candidate solutions before executing them. This is one of the reasons why we consider GMLA-based solutions more suitable for dynamic systems like WSN applications.

The efficiency variation is calculated as follows:

$$\Delta E_f = \left[ \left( \frac{E_{f_{i-1}}}{E_{f_i}} \right) - 1 \right] \times 100 \quad (4)$$

#### 5. Simulation Results

GMLA was evaluated through simulations using the TrueTime simulator<sup>1</sup>. IEEE 802.15.4 was used as the network protocol and the slave density was varied in order to show that GMLA is able to adapt the  $S_p$  in several levels of slave densities. The MAC and physical parameters for the IEEE 802.15.4 network considered in the simulations are represented in Table 2.

**Table 2: IEEE 802.15.4 parameters**

Data Rate	Transmit Power	Receiver Signal Threshold	Pathloss Exponent	Ack Timeout	Retry Limit
250 Kbps	-3 dbm	-48 dbm	3.5	0.864 ms	3

Some of the most relevant parameters used to set the GMLA are represented in Table 3.

**Table 3: GMLA parameters**

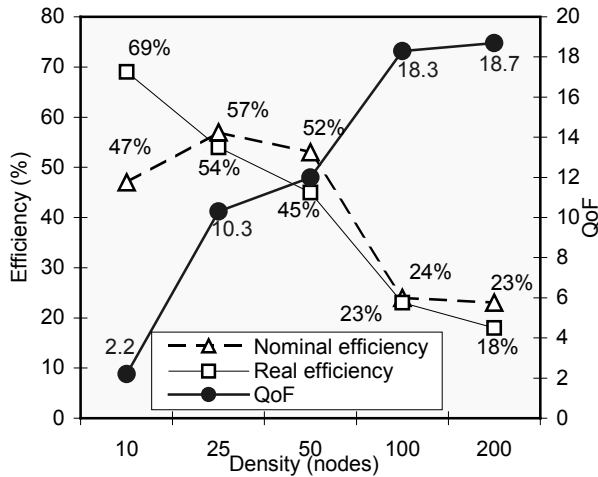
Checkpoint Time	Reposition Rate	Evolution	Population Size	Action Tax
5 rounds	2	15 checkpoints	16	100

Three metrics were considered in our simulations: (i) nominal efficiency: communication efficiency that is based on the master node efficiency estimation; (ii) real efficiency: that is evaluated based on real values; and (iii) Quality of Fusion.

The difference between nominal and real efficiency (Figure 3) demonstrates how much the efficiency is well estimated by master node. This estimation is very important in our approach, as the master node does not have an exact vision of the system (number of slave nodes that are well-functioning).

Figure 3 represents the GMLA behaviour when the slave density is modified. It can be noticed that it tries to maximize  $E_f$ , but it maintains  $Q_{\odot F}$  in a certain level.

<sup>1</sup> <http://www.control.lht.se/truetime/>



**Figure 3. GMLA Simulation Results**

The main target of the proposed approach is to achieve a trade-off between  $E_f$  and  $QoF$ . It can be noticed that  $S_p$  has a great influence on  $E_f$ . If  $S_p$  is low there will have less messages being sent, and  $E_f$  will be maximized (lower load and packets collision will be reduced). However, if  $S_p$  achieves a minimum level,  $E_f$  could be 0 because no message will be sent in the network. When this level of  $S_p$  is achieved our approach detects a low level of  $S_p$  and it tries to increase it.

Table 4 shows a comparison between the IEEE 802.15.4 standard algorithm (in the beaconless mode) and the GMLA proposed approach. Considering that GMLA main target is to maximize the  $E_f$ , it succeeded. However, the  $QoF$  parameter was smaller in all cases. This simulation tests show that is possible to use GMLA approaches in self-organizing WSN's. Considering that in most cases  $QoF$  can be reduced in order to increase the Efficiency ( $E_f$ ) parameter, the main target of this research work was achieved.

**Table 4: IEEE 802.15.4 and GMLA comparison**

Density	IEEE 802.15.4		GMLA	
	$E_f$	$QoF$	$E_f$	$QoF$
10	68%	6.8	69%	2.2
50	16%	32	45%	12
100	19%	19	23%	18.3
200	20%	10.5	18%	18.7

## 6. Final Considerations

In this paper we presented a GMLA approach that self-organizes a WSN data fusion application. Our main goal is to maximize the communication efficiency in a IEEE 802.15.4 in beaconless mode with star topology.

GMLA shows that the Efficiency ( $E_f$ ) parameter can be maximized even for communication scenarios with unknown slave density. However, we consider that the classifiers condition as being just one of the objectives. We are currently working in a GMLA approach that penalizes classifiers that does not achieve a certain  $QoF$  threshold. This way, it will become possible to make a more realistic trade-off between quality of fusion and communication efficiency.

## References

- [1] Q. Wu, N.S.V. Rao, J. Barhen, S.S. Iyergan, V.K. Vaishnavi, H. Qi, K. Chakrabarty, On Computing Mobile Agent Routes for Data Fusion in Distributed Sensor Networks, IEEE Transactions on Knowledge and Data Engineering, Vol 16, No. 6, 2004.
- [2] A.A. Somasundara, A. Rammoorthy, M.B. Srivastava, Mobile Element Scheduling with Dynamic Deadlines, IEEE Transactions on Mobile Computing, Vol. 6, No.4, 2007.
- [3] J.A. Stankovic, T.F. Abdelzaher, C. Lu, L. Sha, J. Hou, Real-Time Communication and Coordination in Embedded Sensor Networks. Proceedings of the IEEE, Vol. 91, No. 7, 2003.
- [4] R. Niu, P.K. Vashney, Q. Cheng, Distributed Detection in a Large Wireless Sensor Network, Science Direct, Information Fusion 7, 2006.
- [5] D. Goldberg, Genetic Algorithm in Search, Optimization and Machine Learning, Addison-Wesley, 1989.
- [6] S. Patil, S. Das, A. Nasipuri, Serial Data Fusion Using Space-Filling Curves in Wireless Sensor Networks, IEEE SECON 2004, 2004.
- [7] K. Kang, H. S. Son, J. A. Stankovic, Managing Deadline Miss Ratio and Sensor Data Freshness in Real-Time Databases, IEEE Transactions on Knowledge and Data Engineering, vol. 16, no. 10, 2004.
- [8] K. Morita, K. Watanabe, N. Hayashibara, T. Enokido, M. Takanizawa, Efficient Data Transmission in a Lossy and Resource Limited Wireless Sensor-Actuator Network, Proceedings of the 10<sup>th</sup> IEEE International Symposium on Object and Component-Oriented Real-Time Distributed Computing (ISORC'07), 2007.
- [9] T. Yan, T. He, J.A. Stankovic, Differentiated Surveillance for Sensor Networks, Proceedings of 1<sup>st</sup> International Conference on Embedded Networked Sensor Systems, 2003.
- [10] IEEE 802.15.4 standard specification, available at <http://www.ieee802.org/15/pub/TG4.html>, 2008.
- [11] A. Koubaa, M. Alves, E. Tovar, A comprehensive Simulation Study of Slotted CSMA/CA for IEEE 802.15.4 Wireless Sensor Networks, IEEE WFCSS, 2006.